DOCUMENT RESUME

ED 139 049 CS 501 695

AUTHOR Monge, Peter R.
TITLE Multivariate Multiple Regression in Communication Research.
PUB DATE Apr 77
NOTE 80p.; Paper presented at the Multivariate Analysis in Communication Research Conference (Pacific Grove, California, April 3-6, 1977)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Communication (Thought Transfer); *Educational Research; Error Patterns; *Hypothesis Testing; Information Theory; Models; *Multiple Regression Analysis; *Predictive Measurement; Predictor Variables; *Statistical Analysis

ABSTRACT

This paper examines the multivariate multiple regression model and explores its applicability to communication research. The first section discusses partitioning of a data matrix, as a heuristic device for distinguishing among alternative regression models. Section two examines various aspects of univariate multiple regression, including the form of the model, the use of the model for descriptive purposes, the properties of estimators, the assumptions of the model, procedures for parameter estimation, procedures for hypothesis testing, interpretation of regression coefficients, and the need for reestimation. The third section offers a parallel presentation of aspects of multivariate multiple regression. A final section considers tests of assumptions, techniques for coping with failed assumptions, the relation of multivariate multiple regression to other multivariate procedures, and the advantages and disadvantages of multivariate multiple regression. (Author/AA)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED139049

Multivariate Multiple Regression in Communication Research

Peter R. Monge

San Jose State University

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Peter R. Monge

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

Running head: Multivariate Multiple Regression

Paper presented at the conference on Multivariate Techniques in Communication Research, Asilomar Conference Grounds, Pacific Grove, California, April 3 - 6, 1977.

CS 501 695

Contents

I. Introduction 2

A. Notation 4

B. Partitioning the Date Matrix for Regression 4

II. Univariate Multiple Regression 8

A. The Univariate Multiple Regression Model 8

B. Description 11

C. Parameter Estimation 17

D. Assumptions of the Linear Regression Model 19

E. Estimating Univariate Regression Parameters 22

F. Hypothesis Testing 26

G. Interpretation of Regression Parameters 30

H. Reestimation of Parameters 40

III. Multivariate Multiple Regression 42

A. The Multivariate Multiple Regression Model 42

B. Estimation of Multivariate Regression Parameters . . 44

C. Testing Multivariate Hypotheses 48

D. Reestimation of Multivariate Parameters 55

IV. Issues and Implications 55

A. Algorithms for Selecting Predictor Variables 55

B. Testing Model Assumptions 61

C. Coping with Failed Assumptions 64

D. Comparisons with Related Techniques 66

E. Advantages and Limitations to MMR 67

F. Conclusion 68

References 69

Tables 71-75

Multivariate Multiple Regression in Communication Research

> Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise. (Tukey, 1962, p. 13)

The scientific study of human communication, like any other science, is fundamentally concerned with establishing laws of relations among the variables that constitute its subject matter. While formulating laws is, of course, a theoretical endeavor, establishing them is an empirical enterprise. When a communication scientist undertakes research to establish a law, he must choose a statistical technique from among the large repertoire available which in his judgment is the one best suited to enable him to derive meaningful conclusions. This paper will present one alternative, multivariate multiple regression (MMR) and its univariate counterpart, which in my opinion is often ideally suited to this task. In fact, as Blalock (1964) says, "It is the regression coefficients which give us the laws of science" (p. 51).

It is not my intention, however, to argue for the superiority of MMR over other analytic techniques. Rather, it seems important that the communication researchers understand the technique and its assumptions so that, as with all statistical procedures, he will

have a rational basis for selecting it when it is best suited to his needs. In this regard, the assumptions are most critical for here is where the researcher must compare the nature and assumptions of the theory being tested (i.e., the proposed law) with the assumptions of the analytic technique. Should the theoretical and analytic assumptions fail to correspond, then the analysis under one set of assumptions of data gathered under the other set is bound to be in error.

The purpose of this paper, then, is to examine the multivariate multiple regression model and to explore its applicability to the domain of communication inquiry. We will begin with a prelude on partitioning a data matrix as a heuristic device for distinguishing among alternative regression models. In the second section we will discuss the major aspects of univariate multiple regression: (a) the form of the model, (b) using the model for descriptive purposes, (c) estimators and their properties, (d) assumptions of the model, (e) procedures for parameter estimation, (f) hypothesis testing procedures, (g) interpretation of regression coefficients, and (h) reestimation. The third section will provide a parallel presentation for multivariate multiple regression. In the final section, we will turn to several additional topics. These will include: (a) tests of assumptions, (b) coping with failed assumptions, (c) the relation of MMR to other multivariate procedures, and (d) the advantages and disadvantages of MMR.

Notation

In terms of notation it may be helpful at the outset to make the conventions we will utilize explicit. For parameters, i.e., characteristics of populations, we shall use upper and lower case Greek letters, e.g., $\underline{\Sigma}$, σ^2, $\underline{\beta}$, μ. For estimators of parameters from sample data, we shall use the same Greek letters as the corresponding parameter augmented by a caret, ^, above each symbol, e.g., $\hat{\beta}$ as an estimator of β. For statistics, i.e., characteristics of samples, and for computational forms, we will use English equivalents (or alternatives, if necessary) to the Greek symbols, e.g., $\underline{S}$ for $\underline{\Sigma}$, $\underline{b}$ for $\underline{\beta}$, etc. Vectors will be identified by lower case letters, matrices by upper case, and both will always be underscored to distinguish them from scalars and other data representations; thus, β is a single regression coefficient, $\underline{\beta}$ is a vector of coefficients and $\underline{B}$ is a matrix of coefficients.

Partitioning the Data Matrix for Regression

Virtually every communication researcher has gathered data for several different variables on some number of subjects. The most traditional way to prepare these data for analysis is to arrange them into a subjects by variables data matrix. In fact, virtually all standard computerized statistical packages require that data be prepared in this way. Here the deck of punched cards may be considered the data matrix, where each card corresponds to the set

of observations (scores or measurements) on a single subject, i.e., a row of the data matrix. Since this format is generally familiar to researchers, let us examine various partitionings of this matrix in order to provide an overview of alternative linear regression models.

Suppose that we have a data matrix, $\underset{N \times M}{A}$, which consists of measurements for N subjects (i = 1, 2, . . ., N) on M variables (j = 1, 2, . . ., M). Our matrix is then of order N x M. Let us assume that we are interested in how the values of some variables can be determined from knowledge about the values of other variables; regression analysis is appropriate for examining this kind of dependency. Some variables in the matrix will be identified as predictor variables and others as criterion variables. We can partition the data matrix into two submatrices, one for predictor variables and one for criterion variables, and examine the interrelations among the two submatrices.

According to convention, we will label the predictor variables the X variables and the criterion variables the Y variables. We will label our criterion submatrix, $\underset{N \times P}{Y}$, and our predictor submatrix, $\underset{N \times K}{X}$. It is possible to have one or more of each kind of variable, i.e., one or more criterion variables and one or more predictor variables, which provide four possible combinations or regression models. The first situation occurs if we have a data matrix with

only two variables. Then we would partition $\underset{NxM}{A}$ (M = 2) as follows (where M = P + K = 1 + 1 = 2)

$$(1) \qquad \underset{Nx2}{A} = \left[\underset{NxP}{Y} \,\vdots\, \underset{NxK}{X}\right] = \left[\underset{Nx1}{Y} \,\vdots\, \underset{Nx1}{X}\right] = \left[\underset{Nx1}{y} \,\vdots\, \underset{Nx1}{x}\right]$$

With one criterion variable and one predictor variable we have the data partition for univariate simple regression. Note that since there is only one criterion and one predictor variable, that the submatrices are really vectors (and $\underset{NxM}{A}$ is N x 2).

Now assume that M is greater than two. We must decide how to partition the matrix. If we choose one variable (the first, for convenience) as the criterion variable, so that P = 1 and treat all the remaining variables as predictor variables, ($K \geq 2$), then we would partition $\underset{NxM}{A}$ ($M \geq 3$) as follows

$$(2) \qquad \underset{NxM}{A} = \left[\underset{NxP}{Y} \,\vdots\, \underset{NxK}{X}\right] = \left[\underset{Nx1}{y} \,\vdots\, \underset{NxK}{x}\right]$$

and call it the data partition for univariate multiple regression. There is a single vector of criterion scores but a matrix of two or more predictor scores (one score for each subject on each predictor variable).

Now consider what would happen if we were to partition $\underset{NxM}{A}$ so that there were two or more criterion variables ($P \geq 2$) but only one predictor variable (K = 1). $\underset{NxM}{A}$ would look like

$$(3) \qquad \underset{NxM}{A} = \left[\underset{NxP}{Y} \;\vdots\; \underset{Nx1}{x}\right] \quad (M = P + K = P + 1 \geq 3)$$

This data partition can still be treated as a regression problem. Since there are several criterion variables but only one predictor variable it would be the appropriate data setup for a multivariate simple regression analysis.

Finally, assume that the researcher is interested in examining the regression of two or more criterion variables ($p \geq 2$) on two or more predictor variables ($K \geq 2$). Then,

$$(4) \qquad \underset{NxM}{A} = \left[\underset{NxP}{Y} \;\vdots\; \underset{Nxk}{X}\right] \quad (M = P + k \geq 4)$$

This partitioning of the data matrix into two submatrices is appropriate for analysis under the multivariate multiple regression model.

During the remainder of this paper we will work with the criterion and predictor submatrices, $\underset{NxP}{Y}$ and $\underset{NxK}{X}$, rather than the full data matrix. Since it is helpful in sorting out the differences for the various forms of analysis, the results of this section are summarized in Table 1. This data partitioning will also be useful when it comes to comparing the regression procedures with related techniques.

Table 1 about here

While we will assume at the outset that the reader is familiar with standard simple and multiple regression techniques, a presentation of the univariate model will help to set the stage for our discussion of the multivariate case. To that task we will now turn.

Univariate Multiple Regression

The Univariate Multiple Regression Model

The classical univariate linear multiple regression model can be given by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} + \varepsilon_i \quad (5)$$

which shows the relationship between two or more (K) predictor or independent variables (X) and a criterion or dependent variable (Y) all measured simultaneously on the subject. The model is called linear (or linear in the parameters) because the effects of the various predictor variables are treated as additive, i.e., Y_i is composed of a linear combination of regression parameters. The regression parameters ($\beta_0 \ldots \beta_k$) are the population partial regression coefficients or weights which are determined from the sample data and used to optimally predict Y_i. β_0 is a scaling constant which absorbs the differences in the scales used to measure the Y and X variables. The ε_i represents the error term, i.e., the extent to which the model fails to predict the criterion

scores, Y_i. This can be seen by rewriting (5) as,

$$\varepsilon_i = Y_i - (\beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik}) \tag{6}$$

Since data are gathered on each subject (i = 1 . . . N) in the sample, there are N equations of the form depicted in equation (5) above

$$\begin{aligned} y_1 &= (1)\beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + \ldots \beta_k X_{1k} + \varepsilon_1 \\ y_2 &= (1)\beta_0 + \beta_1 X_{21} + \beta_2 X_{22} + \ldots \beta_k X_{2k} + \varepsilon_2 \\ &\vdots \\ y_N &= (1)\beta_0 + \beta_1 X_{N1} + \beta_2 X_{N2} + \ldots \beta_k X_{Nk} + \varepsilon_N \end{aligned} \tag{7}$$

These equations may be represented in matrix form as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots & X_{1k} \\ 1 & X_{21} & X_{22} & \ldots & X_{2k} \\ 1 & X_{N1} & X_{N2} & \ldots & X_{Nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{8}$$

or

$$\underset{(N\times 1)}{\underline{y}} = \underset{(N\times q)}{\underline{X}} \; \underset{(q\times 1)}{\underline{\beta}} + \underset{(N\times 1)}{\underline{\varepsilon}} \tag{9}$$

$\underline{y}$ is the Nx1 vector of criterion or dependent scores, one for each subject in the sample. $\underline{X}$ is the Nxq model matrix which contains an initial vector of unities and the data for the k predictor or

independent variables ($q = 1 + k$). $\underline{\beta}$ is the $q \times 1$ vector of partial regression coefficients and $\underline{\varepsilon}$ is the $N \times 1$ vector of error terms.

The model given in (9) is often called the raw or raw score form of the linear regression of $\underline{y}$ on $\underline{x}$. Two other forms of the model are also possible and frequently encountered in regression work. One alternative following Timm (1975, p. 270), is called the reparameterized or deviation model. In this model all predictor scores are measured as deviations about their respective means. Thus, the regression equation would be

$$(10) \qquad y_i = \beta_0 + \sum_{j=1}^{k} \beta_j \bar{X}_j + \sum_{j=1}^{k} \beta_j (X_{ij} - \bar{X}_j) + \varepsilon_i$$

which can be represented in matrix notation as

$$(11) \qquad \underset{N\times 1}{\underline{y}} = \underset{N\times q}{\underline{X}_d} \; \underset{q\times 1}{\underline{\eta}} + \underset{N\times 1}{\underline{\varepsilon}}$$

where $q = k + 1$, and $\underline{X}_d$ is the model matrix of deviation scores, and $\underline{\eta}$ (i.e., eta) is the vector of regression coefficients for the reparameterized model.

The other alternative is to standardize the elements of both $\underline{y}$ and $\underline{X}$, which is accomplished by dividing each deviation score by the standard deviation of that variable. The standardized regression model is

(12)
$$\begin{bmatrix} y_{1z} \\ y_{1z} \\ \vdots \\ y_{Nz} \end{bmatrix} = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1k} \\ Z_{21} & Z_{22} & \cdots & Z_{2k} \\ \vdots & \vdots & \vdots\ \vdots & \vdots \\ Z_{N1} & Z_{N2} & \cdots & Z_{Nk} \end{bmatrix} \begin{bmatrix} \gamma_{1z} \\ \gamma_{2z} \\ \vdots \\ \gamma_{Nz} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

or

(13)
$$\underset{Nx1}{\underline{y}_z} = \underset{Nxk}{\underline{Z}} \quad \underset{kx1}{\underline{\gamma}_z} + \underset{Nx1}{\underline{\varepsilon}}$$

Note that the orders of the data matrix $\underline{Z}$ and the regression vector $\underline{\gamma}_z$ in equation (13) are k rather than q = k + 1. Since all variables are standardized to the same scale, there is no longer the necessity for a scaling term; hence, there is no initial vector of unities in $\underline{Z}$, and no γ_{z_0} in $\underline{\gamma}_z$.

Description.

Suppose that a communication researcher has a set of data which he wishes simply to describe. Multiple regression procedures may be used straightforwardly in this case as a descriptive device. The data will be divided into a criterion variable and one or more predictor variables. Using English letters to indicate that regression coefficients are to be calculated from sample data for descriptive purposes only, the regression equation specifying the relation among the variables is given by

(14)
$$\underset{Nx1}{\underline{y}} = \underset{Nxq}{\underline{X}} \quad \underset{qx1}{\underline{b}} + \underset{Nx1}{\underline{e}}$$

where $\underline{y}$ is an N x 1 column vector of observations on the criterion score (sometimes called the regressand), X is the N x q (q = 1 + k) data matrix of observed predictor variables (sometimes called regressors) augmented by an initial vector of unities, $\underline{b}$ is the unknown q x 1 vector of regression coefficients and $\underline{e}$ is the unknown N x 1 vector of errors. The problem is one of determining the unknown regression coefficients and calculating the error components of the equation.

By manipulation of (14) we can create an equation which will give us a solution for the regression parameters. In order to obtain the best possible solution, we would like our regression coefficients when multiplying the $\underline{X}$ scores to reproduce the $\underline{y}$ scores as exactly as possible. Another way to state this is that the error component of the model, the difference between observed and calculated Y scores, or more precisely the error sum of squares, will be at a minimum. First we calculate the error sums of squares, $\underline{e}'\underline{e}$

(15)
$$\underline{e} = \underline{y} - \underline{Xb} \qquad \text{rearranging (14)}$$
$$\underline{e}'\underline{e} = (\underline{y} - \underline{Xb})' (\underline{y} - \underline{Xb}) \qquad \text{multiplying}$$
$$= \underline{y}'\underline{y} - \underline{yXb} - \underline{b}'\underline{X}'\underline{y} + \underline{b}'\underline{X}'\underline{Xb} \qquad \text{collecting terms}$$
$$= \underline{y}'\underline{y} - 2\underline{b}'\underline{X}'\underline{y} + \underline{b}'\underline{X}'\underline{Xb}$$

Since we are interested in finding the $\underline{b}$ which gives the smallest error sums of squares, let us differentiate these error sums of

squares with respect to $\underline{b}$

$$(16) \qquad \frac{\partial(\underline{e}'\underline{e})}{\partial\,\underline{b}} = -2\underline{X}'\underline{y} + 2\underline{X}'\underline{X}\underline{b}$$

This solution to $\partial(\underline{e}'\underline{e})/\partial\underline{b}$, when set equal to 0, gives the "normal equations"

$$(17) \qquad \underline{X}'\underline{X}\underline{b} = \underline{X}'\underline{y}$$

Here $\underline{X}'\underline{X}$ is the sums of squares and cross products for the predictors, $\underline{b}$ is the vector of regression coefficients, and $\underline{X}'\underline{y}$ is the sum of squares and cross products for the predictor and criterion variables.

Simple algebraic manipulation of (17) gives us the solution for the regression coefficients we seek.

$$(18) \qquad \underline{b} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$$

For those familiar with the calculation of simple regression coefficients in summation notation it might be useful to point out that in the bivariate case (18) is equivalent to

$$(19) \qquad b_{yx} = \frac{\Sigma x y}{\Sigma x^2}$$

which is the ratio of variation between x and y to variation in x alone.

Having now obtained $\underline{b}$ it is possible to insert the values of b into the equation to <u>calculate</u> or <u>predict</u> the values of $\underline{y}$. Since we

already possess the observed predictor scores, $\underline{X}$, the calculated scores, $\hat{\underline{y}}$, are

$$\hat{\underline{y}} = \underline{Xb} \tag{20}$$

If we now compare our calculated criterion values, $\hat{\underline{y}}$, with the observed criterion values, $\underline{y}$, we will discover the extent to which our predictor variables, $\underline{X}$, when multiplied by the best possible regression coefficients, $\underline{b}$, accurately reproduce the observed scores, $\underline{y}$. As a measure of our failure to correctly reproduce the observed scores from our predictor variables and the regression coefficients we construct an equation for "residuals."

$$\hat{\underline{e}} = \underline{y} - \underline{Xb} = \underline{y} - \hat{\underline{y}} \qquad \text{since } \hat{\underline{y}} = \underline{Xb} \tag{21}$$

where $\hat{\underline{e}}$ is the vector of residuals obtained by comparing the observed $\underline{y}$ scores with the calculated or predicted scores, $\hat{\underline{y}}$.

Having obtained our best weights for the set of predictors it is useful to determine how well the prediction model, $\underline{Xb}$, "fits" the observation vector, y. The procedure, following Goldberger (1964, p. 159), is to partition or decompose the variance of $\underline{y}$ into its component parts. The parts of the decomposition can then be developed

into a measure of the goodness of fit, R^2. First, we obtain the sums of squares for residuals.

(22) $\hat{\underline{e}}'\hat{\underline{e}} = (\underline{y} - \underline{Xb})' (\underline{y} - \underline{Xb})$ from (21)

$= \underline{y}'\underline{y} - \underline{y}'\underline{Xb} - \underline{b}'\underline{X}'\underline{y} + \underline{b}'\underline{X}'\underline{X}\,\underline{b}$ multiplying

$= \underline{y}'\underline{y} - 2\underline{b}'\underline{X}'\underline{y} + \underline{bX}'\underline{X}\,(\underline{XX})^{-1}\,\underline{X}'\underline{y}$ combining terms and from (18)

$\hat{\underline{e}}'\hat{\underline{e}} = \underline{y}'\underline{y} - \underline{b}'\underline{X}'\underline{y}$ since $\underline{X}'\underline{X}\,(\underline{XX})^{-1} = \underline{I}$

Also, from equation (20) we can obtain the sums of squares total for the calculated scores, $\hat{\underline{y}}$.

(23) $\hat{\underline{y}}'\hat{\underline{y}} = (\underline{Xb})' (\underline{Xb})$ squaring (20)

$= \underline{b}'\underline{X}'\underline{Xb}$ rearranging

$= \underline{b}'\underline{X}'\underline{y}$ since $\underline{Xb} = \underline{y}$

By substituting (23) in (22) and rearranging terms, we obtain sums of squares total for the observed $\underline{y}$ scores.

(24) $\underline{y}'\underline{y} = \hat{\underline{y}}'\hat{\underline{y}} + \hat{\underline{e}}'\hat{\underline{e}}$

This provides the fundamental partition for linear regression. The partition states that total variation in observed $\underline{y}$, can be decomposed

into two components: (1) sums of squares for predicted $\hat{y}$, also called sums of squares for regression, SS_R, and (2) sums of squares error, SS_E. If we treat the sums of squares as deviations from their respective means, then we have

$$(25) \qquad SS_T = SS_R + SS_E$$

As a measure of the "goodness of fit," the sample coefficient of multiple determination, R^2, is calculated.

$$(26) \qquad R^2 = 1 - \frac{SS_E}{SS_T} = 1 - \frac{\Sigma\hat{e}^2}{\Sigma(y-\bar{y})^2} = \frac{SS_R}{SS_T}$$

which varies between 0 and 1. As Goldberger (1964) says, "When the fit is perfect, the least squares plane passes through every observed y, every $\hat{e} = 0$, so $R^2 = 1$. At the other extreme $b_1 = \ldots = b_k = 0$, $b_0 = \bar{y}$, the plane is horizontal at $\bar{y}$, every $\hat{e} = y - \bar{y}$, so $R^2 = 0$." (p. 160). R, the square root of R^2, is called the sample multiple correlation coefficient.

The results of this section indicate that regression analysis may be used for purely descriptive purposes. Regression coefficients may be obtained and the adequacy of the predictor variables may be determined as proportions of variance accounted for or as goodness

of fit between observed and predicted scores.

Communication researchers are rarely interested in simple description. Typically, we draw samples and wish to make inferences back to the population from which the sample came. In regression analysis, this inference will typically encompass two different but related processes: (1) estimation of population parameters (regression coefficients) from sample data, and (2) hypothesis tests regarding the parameters.

Parameter Estimation

The process of making an inference about the value of a population parameter, θ, from a sample statistic is called estimation. An estimator, $\hat{\theta}$, is a function or formula which tells how to combine sample observations in order to make the estimate about the parameter. An estimate is the value (scalar) obtained for any given sample from an estimator formula. According to Kmenta (1971, p. 9) characteristics of estimators are derived by examing their sampling distributions. The parameters typically estimated in regression analysis are the regression coefficients, $\beta_0 \ldots \beta_k$, i.e., the intercept and slope values of the regression equation.

As many econometricians point out (e.g., Kmenta, 1971, Goldberger, 1964) it is important that estimators possess certain properties:

(1) An estimator should be unbiased. $\hat{\theta}$ is an unbiased estimator of θ if $E(\hat{\theta}) = \theta$. This definition states that on the average, the estimator is correct, which implies that the mean of the sampling distribution of the estimator equals the population parameter.

(2) Unbiased estimators should be efficient. An $\hat{\theta}$ is efficient if it has minimum variance among the class of unbiased estimators.

(3) An estimator should also be consistent. This implies that as sample size gets larger (i.e., approaches the population size), the estimator provides better estimates. An alternative way to state this is that the sampling distribution of the estimator tends to concentrate on the true value of the parameter, i.e., it becomes less biased and smaller variance.

(4) Finally, an estimator that is consistent is Best Asymptotic Normal (BAN) ". . . if the asymptotic distribution of $\sqrt{N}(\hat{\theta}_N - \theta)$ is normal with mean 0 and variance $\sigma^2(\theta)$ has the least possible value." (Timm, 1975, p. 151)

Many econometricians call an estimator, $\hat{\theta}$, that is an unbiased, minimum variance linear function of sample observations a Best

Linear Unbiased Estimator (BLUE) (See, e.g., Goldberger, 1964, p. 127). Technically, bias, minimum variance and BLUE are referred to as finite (small) sample properties, while consistency and BAN are asymptotic (large) sample properties. For an illuminating analogy between properties of estimators and shooting at a bull's eye target with a rifle, see (Kmenta, 1971, pp. 13-14 and 168). Although it will not be proved in this paper, it will come as no surprise to many readers, that the least squares procedures which were discussed earlier for determining regression coefficients for descriptive purpose, turn out to be, under certain assumptions, BLUE of population parameters. We now turn to those assumptions.

Assumptions of the Linear Regression Model

In order to make parameter estimates on the basis of sample data it is necessary to make a number of assumptions about the population. If these assumptions are warranted, then statistical theory regarding sampling distributions and properties of estimators can be used to formulate inferences about the parameters.

The assumptions of the classical linear regression model can be summarized by several equations which will be briefly discussed in this section.

Assumption (1) is

$$\underline{y} = \underline{X\beta} + \underline{\varepsilon} \tag{27}$$

which specifies the functional form of the relationship in the population. It states that the observed scores, the y_i, are linearly dependent upon the X_{ij} scores and the disturbance or error terms, $\underline{\varepsilon}_i$.

Assumption (2) is

(28) $\quad E(\underline{\varepsilon}) = \underline{0} \quad$ or, alternatively $\quad E(\underline{y}) = \underline{X\beta}$

This assumption states that each disturbance term has an expected value of zero. The two forms of the assumption are equivalent because if $\underline{y} = \underline{X\beta} + \underline{\varepsilon}$ and $E(\underline{y}) = \underline{X\beta}$, then it must follow that $E(\underline{\varepsilon}) = \underline{0}$.

Assumption (3) also may be written in two forms:

(29) $\quad V(\underline{\varepsilon}) = E(\underline{\varepsilon}'\underline{\varepsilon}) = \sigma^2\underline{I} \quad$ or $\quad V(\underline{y}) = \sigma^2\underline{I}$

The variance of $\underline{\varepsilon}$ equals the variance of $\underline{y}$ because $\underline{\varepsilon}$ and $\underline{y}$ are separated only by a constant, which does not affect the variance. This equation states that the expected value for the sums of squares for the disturbances will equal a constant variance times the identity matrix. This really encompasses two assumptions. First, assumption (3) specifies that for all values, X_{ij}, of any given predictor variable, X_j, the variances will be constant, i.e., they

will be homoskedastic. The assumption of homoskedasticity is sometimes written as

(29a) $$E(\varepsilon_i^2) = \sigma^2$$

Violation of the assumption, i.e., $E(\varepsilon_i^2) \neq \sigma^2$ is referred to as heteroskedasticity. Second, since an identity matrix has zeroes in the off-diagonal position, (3) implies that the errors are uncorrelated, which can also be written as

(29b) $$E(\varepsilon_i \varepsilon_j) = 0 \text{ for all } i \neq j.$$

This assumption is often referred to as one of the nonautoregression or nonautoregressive disturbances.

Assumption (4) stipulates that

(30) $\underline{X}$ is an N x q matrix which is fixed in repeated samples.

This assumption implies a nonstochastic X which further implies that $\underline{X}$ and $\underline{\varepsilon}$ are independent.

Assumption (5) is that the

(31) Rank of $\underline{X} = q \leq N$

which indicates that there are more subjects than variables and that no exact linear combinations exist among the predictor variables. This latter statement, that there is no exact correlation among the predictors, is often referred to as lack of multicollinearity. In practice we are often more concerned with high degrees of multicollinearity than with perfect multicollinearity.

If we add one additional assumption, (6) that the ε_i are normally distributed, to the previous five, the model becomes the classical normal linear regression model. Assumptions 1 - 3 and 6 may be compactly summarized for this model as

(32) $\underline{y} \sim N(\underline{X\beta}, \sigma^2\underline{I})$ or, alternatively $\underline{\varepsilon} \sim N(\underline{0},\underline{\Sigma})$, where

$$\underline{\Sigma} = \sigma^2\underline{I}$$

This format for synthesizing assumptions 1 - 3 and normality may be read as "Y is a normally distributed random variable with expectation (mean) equal to $\underline{X\beta}$, and variance equal to $\sigma^2\underline{I}$," or alternatively, "The errors are normally distributed with zero mean and equal variance, $\underline{\Sigma}$." It is also often assumed that the X_{ij} scores are measured on scales that have at least ordinal properties, but that is not required by either model.

Estimating Univariate Regression Parameters

Having stated a model of the form

$$\underline{y} = \underline{X}\beta + \underline{\varepsilon}$$

and specified a number of assumptions about that model, how can we obtain BLUE estimates of the regression parameters? If the assumptions are valid, it turns out that the least squares procedures that were utilized for descriptive purposes, may also be used for inferential purposes. Using the carot (^) to indicate that sample data are being used to estimate the population regression parameters, equation (9) is rewritten as

$$(33) \qquad \underset{Nx1}{\underline{y}} = \underset{Nxq}{\underline{X}} \quad \underset{qx1}{\hat{\underline{\beta}}} + \underset{Nx1}{\hat{\underline{\varepsilon}}}$$

where the details of the vectors and matrices are as previously specified. Equation (33) may be manipulated in identical fashion to the way equation (18) was derived from equation (14). These operations lead to the least squares estimate of the population regression coefficients (See Finn, 1974, p. 97, for a demonstration without using calculus that $\hat{\beta}$ is a minimum sum of squared errors estimate of β.).

$$(34) \qquad \hat{\underline{\beta}} = (\underline{X}\underline{X})^{-1}\,\underline{X}\,\underline{y}$$

$\hat{\underline{\beta}}$ is an unbiased estimator of $\underline{\beta}$ because $E(\hat{\underline{\beta}}) = \underline{\beta}$. This can be shown

fairly easily.

$$
\begin{aligned}
(35) \qquad E(\hat{\underline{\beta}}) &= E\left[(\underline{X}'\underline{X})^{-1}\ \underline{X}'\underline{y}\right] && \text{taking expertations of (34)} \\
&= (\underline{X}'\underline{X})^{-1}\ \underline{X}\ E\ (\underline{y}) && \text{since } \underline{X} \text{ is constant} \\
&= (\underline{X}'\underline{X})^{-1}\ \underline{X}'\underline{X}\beta && \text{from Assumption 2} \\
& && E(\underline{y}) = \underline{X}\beta \\
&= \underline{\beta} && \text{since } (\underline{X}'\underline{X})^{-1}\ \underline{X}'\underline{X} = \underline{I}
\end{aligned}
$$

$\hat{\underline{\beta}}$ is also efficient, i.e., minimum variance. (See Finn, 1974, p. 99, for proof). And in general, $\hat{\underline{\beta}}$ is a BLUE of $\underline{\beta}$ (see Kmenta, 1971, pp. 209-216).

Since an estimate, though unbiased, will be correct only on the average (i.e., $E(\hat{\theta}) = \theta$), it is important to obtain an estimate of the variability of $\hat{\underline{\beta}}$ over repeated samples (of the same size). This, of course, is equivalent to asking, "What is the variance or standard deviation of the sampling distribution of the estimator, in this case $\hat{\underline{\beta}}$?" The variance of $\hat{\underline{\beta}}$ (see Johnston, 1972, pp. 125, 126) is given by

$$
(36) \qquad V(\hat{\underline{\beta}}) = \sigma^2\ \text{diag}\ (\underline{X}'\underline{X})^{-1}
$$

and the standard deviation of $\hat{\underline{\beta}}$, called the standard error of estimate of $\hat{\underline{\beta}}$ is given by

$$(37) \qquad \sigma(\hat{\underline{\beta}}) = \sigma \text{ diag } (\underline{X}\ \underline{X})^{-1}$$

Thus, for the i^{th} regression coefficient, $\hat{\beta}_j$, the standard error of estimate is

$$(38) \qquad \sigma_{\hat{\beta}_j} = \sigma \text{ diag } (\underline{X}\ \underline{X})_{jj}^{-1}$$

Both the variance of $\hat{\underline{\beta}}$ and the standard error of estimate can be found on the diagonals of their respective matrices.

The conventions regarding reporting of regression results in economics might usefully be incorporated into communication research. As Kmenta, 1971, indicates, "It has become customary to present all these results by writing out the estimated regression equation with the estimated standard errors in parentheses under the respective coefficients. This is followed by the value of R^2" (p. 242). For our development we would write

$$(39) \qquad Y_i = \hat{\beta}_o + \hat{\beta}_1 X_i + \ldots + \hat{\beta}_k X_k + e_i \qquad R^2 = \ldots$$
$$(\sigma_{\hat{\beta}_o})(\sigma_{\hat{\beta}_1}) \quad \ldots \quad (\sigma_{\hat{\beta}_k})$$

The foregoing indicates that the procedures for determining regression coefficients for inferential purposes (parameter estimation) are no different from those employed for descriptive purposes. What does differ is that in the former case a number of critical

assumptions are made about the population, while in the latter case, no assumptions are made. These assumptions, of course, make all the difference. Depending on the extent to which they are invalid, the researcher must either abandon ordinary least squares (OLS) techniques in favor of other alternatives (e.g., two stage least squares, instrumental variables, etc.) or be left in the unenviable position of being able only to describe relations in the obtained data. We will briefly review those alternatives at a later point in the paper. Now, having shown how to obtain parameter estimates, we will briefly review various interpretations of regression coefficients before turning to the important topic of how to test the significance of those estimates.

Interpretation of Regression Coefficients

It might be helpful at this point to be explicit about the interpretation of the various regression coefficients. As already mentioned, the parameter β_o, its estimator counterpart, $\hat{\beta}_o$, and the descriptive sample coefficient, b_o, are all scaling coefficients which are necessitated if $\underline{y}$ and $\underline{X}$ are measured on different scales, i.e., in different units of measurement. These coefficients insure the equality of the right hand and left hand sides of their respective equations.

(40) $$y = \beta_o + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$$

$$\beta_o = y - (\beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon)$$

If the variables are all measured on the same scale, the coefficient will equal zero. Thus, in the standardized model where all variables are converted to the same scale by Z-score transformations there is no γ_0, $\hat{\gamma}_0$ or g_0 coefficients (Remember, γ_0 is the standardized coefficient that corresponds to β_0); β_0, $\hat{\beta}_0$ and b_0 are defined only for the raw and deviation forms of the regression model. Each may also be considered the intercept of the line, plane or hyperplane of the regression equation (depending upon whether it is a simple regression, two variable multiple regression, or greater than two variable multiple regression equation). As an intercept, it is the value of y when $X_1 = 0$ in the one predictor case, when X_1 and $X_2 = 0$ in the two predictor case or $X_1 \ldots X_k = 0$ in the greater than two predictor variable case. It should be pointed out that many texts also call this coefficient α, $\hat{\alpha}$, and a.

The other regression coefficients, $\beta_{1 \ldots k}$, $\hat{\beta}_{1 \ldots k}$, and $b_{1 \ldots k}$ are called partial regression coefficients. They should be interpreted as the change in y per unit change in X_j. This quantity is typically referred to as the slope of the line, plane or hyperplane of the regression of y on X_j. Alternatively, they may be thought of as weights which are used as multipliers for the X_j in order to optimally predict y. They are called partial regression coefficients because each shows the relationship between y and a given X_j partialing out or controlling for the influences

of all other Xs in the equation. Though not always stated, the term regression coefficient is always understood to mean partial regression coefficient. (The only exception to this is the intercept or scaling coefficient, β_0, which is why some authors prefer to give it an entirely different symbol, i.e., α.) As should also be clear shortly, they are sometimes referred to as unstandardized regression coefficients, where the term "partial" is again understood. Finally, since unstandardized regression coefficients are appropriately used only with the raw score or deviation forms of the regression equation, they are sometimes referred to as raw regression coefficients.

It is important to note that changing the order of variables in the equation (e.g., exchanging the first and last variables, X_1 and X_k) will not alter the magnitude of any of the regression coefficients. Adding or deleting variables, on the other hand, will typically change the magnitude of the coefficients, in some cases, quite drastically. This is so because partial regression coefficients are defined for a given set of predictors; change the set and the coefficients also change.

Interpretation of a partial regression coefficient by comparing it to other regression coefficients in an equation is difficult if the variables are measured on different scales. Regression coefficients will differ simply as a function of the difference in scales

on which the respective X_j variables were measured in addition to any contributions they make to regression. To overcome the problem of differences in regression coefficients attributable to scaling it is customary to calculate the standardized partial regression coefficients, $\gamma_{1 \ldots k}$, $\hat{\gamma}_{1 \ldots k}$, and $g_{1 \ldots k}$. A standardized regression coefficient is equivalent to a raw coefficient multiplied by the standard deviation of X_j and divided by the standard deviation of y.

$$\gamma_k = \beta_k \frac{\sigma_k}{\sigma_y} \tag{41}$$

Since all variables are measured on the same scale the standardized coefficients are more directly interpretable. A unit change in y per unit change in X_j will give a regression line at 45^0 from the origin, for each X_j. Multiples and fractions of change in y per unit change in X_j are also the same for each X_j. Therefore, the comparison of the magnitudes of the various γ, $\hat{\gamma}$ and g within an equation is facilitated. As Blalock (1971) indicates, standardized regression coefficients are probably best suited for comparing relationships within equations while unstandardized coefficients are more appropriate for "comparing populations or stating general laws." (p. 145; see also, Tukey, 1954, Cp. 3.)

It should be pointed out that one common convention in some

social sciences is to call the standardized regression coefficient a "beta" weight and to reserve for it the symbol "β." This is not the convention in econometrics and many multivariate texts where Greek letters are reserved for population parameters or estimators. (Finn, 1974, in fact, reverses the convention and uses β for the unstandardized and b for the standardized coefficients. Goldberger's, 1964, usage is such that he directly warns his readers about not confusing "beta" coefficients with the elements of the population coefficient vector $\underline{\beta}$.) Sorting out the variety of conventions of b and β is not always easy and it is the main reason why, following Timm (1975) and others, I have chosen to use an entirely different symbol, γ, for the standardized coefficients. The important thing to remember is not the symbol; rather, remember (1) there are two forms of the slope coefficients, standardized and unstandardized, (2) the interpretation appropriate to each, and (3) that each may be defined as a parameter an estimator or a descriptive statistic.

Hypothesis Testing

Hypothesis testing in regression analysis is the procedure whereby influences about the significance of population regression coefficients (parameters) are made on the basis of sample regression coefficients (statistics). As Finn (1975, p. 134) indicates, this process requires two separate steps: (1) partitioning of variation

and covariation in the criterion variables into components for different (sets of) predictor variables, which consists of hypothesis sums of squares (and cross products) and error or residual sums of squares (and cross products), and (2) a comparison between the hypothesis sums of squares and error sums of squares with one or more test statistics.

In a previous section we partitioned the total variance, SS_T, into sums of squares for regression, SS_R, and those for error, SS_E. Recalling equations (22 - 25), let us now use $\hat{\underline{\varepsilon}}$ as an estimator of $\underline{\varepsilon}$ (instead of using $\underline{e}$ for descriptive purposes) and call this scalar quantity, for convenience, $Q_{\hat{\varepsilon}}$. Thus, the sums of squares (and cross products) due to error are

$$(42) \qquad Q_{\hat{\varepsilon}} = \hat{\underline{\varepsilon}}'\hat{\underline{\varepsilon}} = \underline{y}'\underline{y} - \hat{\underline{\beta}}'\underline{X}\underline{X}\hat{\underline{\beta}} \qquad \text{from (22)}$$

and the sums of squares (and cross products) hypothesis, or as it is often called, sums of squares for regression, are

$$(43) \qquad Q_h = \hat{\underline{y}}'\hat{\underline{y}} = \hat{\underline{\beta}}'\underline{X}'\underline{X}\hat{\underline{\beta}} \qquad \text{from (23)}$$

We can how establish null and alternative hypotheses that will enable us to test whether all of the regression coefficients are jointly significant, i.e., whether the entire vector of coefficients,

$\underline{\beta}$, is significant. Thus,

(44) * $H_o : \underline{\beta} = 0$

with the alternative

(45) $H_A : \underline{\beta} \neq 0$

To test (42) we calculate a standard $\underline{F}$ ratio of the two sums of squares divided by their degrees of freedom, i.e., a ratio of two mean squares, MS_h and $MS_{\hat{\varepsilon}}$.

$$(46) \qquad \underline{F} = \frac{Q_h/(k+1)}{Q_{\hat{\varepsilon}}/(N-k-1)} = \frac{\hat{\underline{\beta}}'\underline{X}'\underline{X}\hat{\underline{\beta}}/(k+1)}{\underline{Y}'\underline{Y} - \hat{\underline{\beta}}'\underline{X}'\underline{X}\hat{\underline{\beta}} \;/(N-k-1)}$$

An ANOVA table (See Table 2) may be constructed to summarize the information for the test with (k+1) and (N-k-1) degrees of freedom.

This particular test, though the most general, is of little practical utility since it tests <u>all</u> of the regression coefficients, including the intercept, $\hat{\beta}_o$. To overcome this problem, it is possible to examine the equivalent test for the standardized vector of regression coefficients, $\underline{Y}_z$, or to partition the vector into <u>two</u> subsets, $\hat{\beta}_o$ and $\hat{\beta}_1 \ldots \hat{\beta}_k$, and examine only the slope coefficients (see Timm, 1975, pp. 273-274; and Goldberger, 1964, pp. 176-177).

Table 2 about here

In the latter case, the hypothesis would be

$$H_o : \underline{\gamma} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{47}$$

where β_o is omitted, and $\underline{\gamma}$ is the vector of reparameterized regression coefficients, $\beta_1 \ldots \beta_k$.

As Timm (1975) indicates, "Testing that all the [slope] coefficients $\beta_1, \ldots, \beta_k$ are equal to 0 is a special case of the more general problem of testing that some subset of the coefficients is 0, with no restrictions on the other elements" (p. 275).

The general procedure is similar to the one just described. Partition $\underline{\beta}$ into two subvectors

$$\underline{\beta} = \begin{bmatrix} \underline{\beta}_o \\ \hline \underline{\beta}_h \end{bmatrix} \tag{48}$$

where $\underline{\beta}_o$ contains the regressors <u>not</u> being tested (including the intercept) and $\underline{\beta}_h$ contains those that are under test by the hypothesis. Thus, the null form and the alternative are

$$H_o : \underline{\beta}_h = \underline{0} \quad \text{and} \quad H_A : \underline{\beta}_h \neq 0 \tag{49}$$

which hypothesize that the variables in the second set which correspond to $\underline{\beta}_h$ significantly contribute to the variance in $\underline{y}$ over and above the regressors in the first set, $\underline{\beta}_o$.

A convenient technique for testing this hypothesis is to formulate two separate regression equations. The first equation contains the first m of the k predictors ($m < k$) which are known or assumed to significantly contribute to criterion variance; the second equation contains all k predictors, i.e., the first m variables not being tested plus the remaining k - m variables which constitute $\underline{\beta}_h$. The respective equations would be

(50a) $$\underset{N\times 1}{\underline{y}} = \underset{N\times(m+1)}{\underline{X}}\ \underset{(m+1)\times 1}{\underline{\beta}} + \underset{N\times 1}{\underline{\varepsilon}} \qquad \text{and}$$

(50b) $$\underset{N\times 1}{\underline{y}} = \underset{N\times q}{\underline{X}}\ \underset{q\times 1}{\underline{\beta}} + \underset{N\times 1}{\underline{\varepsilon}} \qquad \text{where } q = k + 1$$

These two equations can, of course, be represented by their respective measures of "goodness of fit," R_o^2 for the reduced set of regressors, $\underline{\beta}_o$, and R_h^2 for the full set of regressors, $\underline{\beta}_o + \underline{\beta}_h = \underline{\beta}$. It is then possible to construct an $\underline{F}$ statistic that will enable us to determine the significance of the $\underline{\beta}_h$ vector over and above the $\underline{\beta}_o$ vector.

(51) $$\underline{F} = \frac{R_k^2 - R_m^2}{1 - R_k^2} \cdot \frac{N - k - 1}{K - m}$$

which has (k - m) and (N - k - 1) degrees of freedom, m equal to the number of X variables in the reduced vector $\underline{\beta}_o$, and k equal to the full set of X variables in $\underline{\beta}$. If $\underline{F}$ is significant at $P \leq \alpha$, then H_o is rejected and it is concluded that some of the regressors in $\underline{\beta}_h$ do contribute significantly to variation in $\underline{y}$. The squared partial multiple coefficient of determination is given by

$$(52) \qquad R^2_{Y\,(m+1,\ldots,k)\cdot(1,\ldots,m)} = \frac{R^2_k - R^2_m}{1 - R^2_m}$$

and indicates the amount of variation in Y which remains after the first m variables have been partialed out of the regression equation.

When $\underline{\beta}_h$ contains only one regression coefficient so that $\underline{\beta}_o = k - 1$ (i.e., $\underline{\beta}_o$ contains all but the last coefficient), equation (51) has a special interpretation which is of considerable importance. Specifically, the equation can be shown to reduce to the partial correlation coefficient (see Timm, 1975, p. 277)

$$(53) \qquad r^2_{Yk\cdot(1,\ldots,k-1)} = \frac{R^2_k - R^2_{k-1}}{1 - R^2_{k-1}}$$

which when examined as an $\underline{F}$ test at 1 and N - k - 1 degrees of freedom shows whether one X variable added to an existing set of predictor

variables contributes significantly to variation in the criterion measure. Thus, we can test whether an individual regression coefficient differs significantly from zero, and thereby contributes significantly to criterion variation. The size of $r^2_{Yk \cdot (1, \ldots, k-1)}$ also provides a measure of the shared variation between the two variables.

There are several other hypotheses that are often worthy of examination; they are listed in Table 3 along with those that have already been discussed.

The formula provided in equation (53) is perfectly adequate for examining the contribution to criterion variance made by the last variable in the equation. In general, however, if we examine more than the final variable we find that the regression coefficients for pairs of predictors are generally correlated. Let us suppose that we wish to obtain individual sequential tests for each of the final $K - m > 1$ predictors, i.e., the last variable, the next to last variable, etc., until we get back to the m^{th} variables. Finn (1974) indicates that

> A series of independent tests if facilitated by transforming the predictor variables to a new set of uncorrelated measures, in a specified order. We shall substitute for predictor X_j in $\underline{X}$ only the linear function or portion of X_j that is uncorrelated with

Table 3 about here

> preceding predictors X_1, X_2, . . . X_{j-1}. That is, we shall find the X_j values that are obtained if we "partial out" or "hold constant" the effects of earlier predictors in the set (p. 137).

There are two important points to note here. The first is that each variable is orthogonalized (made to be uncorrelated) with only those variables that preceed it in the equation. The second is that the process requires that the variables be arranged in a prespecified order.

The process which Finn (1974, pp. 134-144; see also Timm, 1975, Kshirsagar, 1959, and Anderson, 1958) describes produces orthogonal estimates of regression coefficients or semipartial regression coefficients. Each semipartial regression coefficient indicates the relation between Y and X_j controlling for the variance contributed by variables preceding X_j in the equation, but not controlling for the variance of those variables which follow it. The process of sequentially estimating the effect of each predictor variable eliminating those preceding it, is called stepwise elimination.

In general, the r^2s between Y and each X_j will not add up to $R^2_{Y \cdot X_{1 \ldots k}}$ because the r^2s are correlated. The squared semipartial correlation coefficients will, however, add to $R^2_{y \cdot X_{1 \ldots k}}$ since the constraint is imposed that each sequential $r^2_{y\,(X_{j \cdot X_{1 \ldots j-1}})}$

will account only for variance not accounted for by preceding variables. Thus, as Kerlinger and Pedhazur (1973, p. 94) indicate

$$(54) \quad R^2_{y \cdot X_{1 \ldots k}} = r^2_{yX_1} + r^2_{y(X_{2 \cdot 1})} + r^2_{y(X_{3 \cdot X_{1 \ldots 2}})} + \ldots + r^2_{y(X_k \cdot X_{1 \ldots k-1})}$$

which is read, "The total amount of shared variance in y and X is given by the sum of the squared semipartial correlation coefficients, i.e., the squared correlation between Y and X_1 plus the variance between Y and X_2 removing the influence of X_1 from the relation with Y, plus . . . plus the variance between Y and X_k removing the influence of X_1 through X_{k-1} from the relation with Y."

Often a researcher is interested in whether the combined effects of two (or more) variables viewed together significantly contribute to criterion variation over and above the variation accounted for by the variables viewed separately. This interaction between (among) variables can be examined in regression analysis by the inclusion of multiplicative or cross-product terms in the prediction equation. Each interaction term is treated as a new predictor variable in the linear model. The new variable is created by cross multiplying the observed values for each subject on the original variables, i.e., $X_{ij_1} X_{ij_2}$. Since the model is still linear this multiplicative term is added to the model and a regression parameter is estimated for the new term. For example, given two variables, X_1 and X_2, the

linear model which includes the interaction term would be

$$(55) \qquad y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \varepsilon_i$$

To calculate interaction terms it is first necessary to standardize the respective variables prior to computation of the cross-product terms. As Finn (1975) indicates, "the dominance of the interaction by one or another variable due to scaling is avoided. The interaction terms themselves need not be standardized" (p. 85). Once this is accomplished, the OLS estimation and hypothesis testing techniques described earlier can be employed. It should be noted, however, that it is customary to add the multiplicative terms into the equation last, thus testing them first if backward elimination techniques are employed.

Several hypotheses have been identified in this section which a researcher might profitably examine. These do not exhaust the possibilities, however. Table 3 presents those discussed plus several other alternatives which are worthy of consideration. Which test to use depends, of course, upon the theory guiding the research. Perhaps an example will help to illustrate this point.

Attention in communication research tends to focus almost exclusively on slope regression coefficients, i.e., $\beta_i (i=1...k)$ to the exclusion of the intercept coefficient, β_0 (or α). Unless

data have been reparameterized to force the intercept to the origin (or some other level) of the coordinate system, the value of the intercept may also be of considerable theoretical interest. If we modify an example given by Kmenta (1971, pp. 204-205) we can see the importance of the intercept. Suppose that media consumption (i.e., number of hours/week spent watching T.V), Y, is regressed on amount of leisure time (number of hours not devoted to gainful employment or primary occupation), X, and a linear relationship of the form $E(Y) = \beta_o + \beta_1 X_1 + \varepsilon$ is determined. The slope coefficient, β, would be interpreted as the marginal propensity to watch T.V., i.e., the amount of increase in watching T.V. for every unit increase in amount of leisure time. Now examine β_o. Assuming that people will consume <u>at least some</u> T.V. even if they have no leisure time, β_o indicates the minimal or subsistence consumption of T.V., i.e., the amount of T.V. exposure when the value of leisure time is zero. Hypotheses regarding the intercept coefficient could be of significant theoretical value. Refering to Table 3 we might utilize test (1), that $H_o : \hat{\beta}_k = 0$, for the intercept coefficient β_o. If our theory were more sophisticated we might utilize test (2), that $H_o : \hat{\beta}_o = M_k$, for some constant, which is tantamount to asserting that we can predict "subsistence" media consumption at some number greater than zero.

<u>Reestimation</u>

As was pointed out earlier in the paper, estimates of

regression parameters are valid, under the assumptions of the model, only for a given set of regressors. If we change the set of regressors, either by adding or deleting predictor variables, the coefficients for the original equation for which the estimates were obtained will no longer be appropriate. New estimates should be obtained.

It should be clear that the hypothesis testing procedures described in the previous section lead , if significant, to the deletion of variables from the full set of variables included in the original equation. Or, alternatively conceived, the procedures lead to the addition of significant variables over and above those already included in the equation. (It is possible, of course, to test the entire set of variables.) Under either conceptualization, obtaining significant results in hypothesis testing is likely to lead to a different set of variables, a different regression equation, than the one with which the researcher began.

In situations where this occurs, the researcher should reestimate the parameters of the final obtained regression equation. The procedures for reestimation are identical to those for estimation. The final set of variables are estimated by OLS procedures, with the variance of the rejected variables pooled with the error variance. A final estimate of the standard error of estimate should also be obtained. The final equation which

contains the final estimates of regression parameters, the final standard errors of estimate, and the recomputed final R^2 should be the results that are presented in the research report.

Multivariate Multiple Regression

In this section we will extend the findings developed for the univariate model to the multivariate case. Most of the presentation will be by analogy with univariate results.

The Multivariate Multiple Regression Model

Let us assume that a communication researcher has two or more variables that he would like to examine as criterion variables in relation to the same set of predictor variables that we assumed in the univariate case. This is a problem for which multivariate regression analysis is appropriate.

The multivariate multiple regression model is given by

$$(56) \quad \underset{N\times p}{\underline{Y}} = \underset{N\times q}{\underline{X}} \; \underset{q\times p}{\underline{B}} + \underset{N\times p}{\underline{E}}$$

This equation states that a data matrix $\underset{N\times m}{\underline{A}}$ has been partitioned into two submatrices $\underset{N\times p}{\underline{Y}}$ which contains two or more criterion variables and $\underset{N\times q}{\underline{X}}$ which contains all the data for the predictor variables augmented by an initial column of unities. Furthermore, the scores in the $\underline{Y}$ matrix are composed of linear combinations of the X scores, each weighted by a regression coefficient, β_k, and added to the

error term, ε_{ij}. That is,

$$(57) \qquad y_{ij} = \beta_{oj} + \beta_{ij}X_{i1} + \beta_{2j}X_{i2} + \ldots + \beta_{kj}X_{ik} + \varepsilon_{ij}$$

$\underline{E}$, then, is an Nxp matrix of error terms, with one column of errors for each criterion variable. If the model is intended to be a multivariate classical linear regression model we assume

$$(58) \qquad E(\underline{Y}) = \underline{XB} \qquad \text{which is equivalent to } E(\underline{E}) = 0$$

since the expectation of $\underline{Y}$ is $\underline{XB}$, and

$$(59) \qquad V(\underline{Y}) = \underline{I}_N \otimes \underline{\Sigma} \qquad \text{which is equivalent to } V(\underline{E}) = \underline{I}_N \otimes \underline{\Sigma},$$

since $\underline{Y}$ and $\underline{E}$ differ only by a constant, namely $\underline{XB}$. The Kronecker product operator, $\otimes$, which is used to define the variance of $\underline{Y}$, produces in this case a diagonal matrix with diagonal elements equal to $\underline{\Sigma}$ and off-diagonal elements, $\underline{0}$. Thus

$$(60) \qquad V(\underline{Y}) = V(\underline{E}) = \underline{I}_N \otimes \underline{\Sigma} = \begin{bmatrix} \underline{\Sigma} & 0 & \ldots & 0 \\ 0 & \underline{\Sigma} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \underline{\Sigma} \end{bmatrix}$$

These assumptions may be summarized as

$$(61) \qquad \underline{E} \sim N_N (\underline{0}, \underline{I} \otimes \underline{\Sigma})$$

which states that $\underset{Nxp}{Y}$ is a multivariate normal distribution, that the errors have constant variance, and that the errors are uncorrelated. It is also possible to state the model in equivalent mean deviation (reparameterized) and standardized forms. In the former case we have

$$\underset{Nxp}{Y} = \underset{Nxq}{X_d} \quad \underset{qxp}{H} + \underset{Nxp}{E} \tag{62}$$

where eta ($\underline{H}$) is the matrix of partial regression coefficients for the mean-adjusted scores, $\underline{X}_d$. In the latter case we have

$$\underset{Nxp}{Y_z} = \underset{Nxk}{Z} \quad \underset{kxp}{\Gamma_z} + \underset{Nxp}{E} \tag{63}$$

where gamma ($\underline{\Gamma}_z$) is the matrix of standardized partial regression coefficients.

Estimation of Multivariate Regression Parameters

In order to estimate the regression parameters we proceed as in the univariate case. The process is to apply the Gauss-Markoff Theorem (see Timm, 1975, 185-188) to obtain the matrix $\underline{B}$ so that the sum of squared errors is minimized. The theorem states that this can be accomplished by minimizing the trace of the sums of squared errors matrix, $\underline{E}'\underline{E}$. Thus, let S_E equal to the sum of squared errors to be minimized, then

(64) $S_E = Tr\ (\hat{\underline{E}}'\hat{\underline{E}})$ since $Tr\ (\hat{\underline{E}}'\hat{\underline{E}}) = \sum_{i=1}^{k}\sum_{j=1}^{p}\hat{\epsilon}_{ip}^2$

$= TR\ (\underline{Y} - \underline{X}\hat{\underline{B}})'(\underline{Y} - \underline{X}\hat{\underline{B}})$ rearranging (56)

When the partial derivatives of S_E with respect to $\hat{\underline{B}}$ are set equal to zero and solved, the following normal equations are obtained

(65) $$\underline{X}'\underline{X}\hat{\underline{B}} = \underline{X}'\underline{Y}$$

which can be solved for $\hat{\underline{B}}$

(66) $$\hat{\underline{B}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$$

This equation states that under the assumptions of the model the best linear unbiased estimates of the matrix of regression coefficients can be obtained directly from the sums of squares and cross products matrices of the raw data, i.e., from $\underline{X}'\underline{Y}$ and the inverse of $\underline{X}'\underline{X}$. Comparison of equation (66) with equations (18) and (34) indicates that the multivariate solution is completely analagous to its univariate counterpart.

$\hat{\underline{B}}$, of course, consists of a matrix of partial regression coefficients. For each criterion variable there are as many coefficients as there are predictor variables, plus one coefficient for the intercept. Thus, if there are two criterion variables and

three predictor variables, there would be 2 x (3 + 1) = 8 regression coefficients. In general, using the orders of our matrices, there will be (q x p) regression coefficients (remember, q = k predictor variables + 1) for the raw form of the regression model. In the standardized model there would be (p x k) coefficients.

Since $\hat{\underline{B}}$ has been estimated from the data, it is desirable to obtain a measure of the stability of the estimate. Sampling theory specifies that upon repeated sampling from a given population we would obtain a distribution of Bs for samples of a given size, N. That sampling distribution would have both an expected value and a variance (as well as higher order moments). Finn (1973) indicates that "The estimate $\hat{\underline{B}}$, like its univariate counterpart, is unbiased and minimum variance" (p. 113). Its expectation and variance-covariance matrix are

$$(67) \quad E(\hat{\underline{B}}) = E\left[(\underline{X}'\underline{X})^{-1}\,\underline{X}'\underline{Y}\right] \quad \text{from (66)}$$

$$= (\underline{X}'\underline{X})^{-1}\,\underline{X}'E\,(\underline{Y}) \quad \text{since } \underline{X} \text{ is constant}$$

$$= (\underline{X}'\underline{X})^{-1}\,\underline{X}'\underline{X}\underline{B} \quad \text{by substitution from (58)}$$

$$= \hat{\underline{B}} \quad \text{since } (\underline{X}'\underline{X})^{-1}\,\underline{X}'\underline{X} = \underline{I}$$

and

$$(68) \quad V\,(\hat{\underline{B}}) = V\left[(\underline{X}'\underline{X})^{-1}\,\underline{X}'\underline{Y}\right] \quad \text{from (66)}$$

$$= (\underline{X}'\underline{X})^{-1} \otimes \underline{\Sigma} \quad \text{See Finn, p. 114, for proof.}$$

Using equation (64) we can also see that an unbiased (maximum likelihood) estimate of the error variance of $\underline{Y}$, $\hat{\underline{\Sigma}}$, can be obtained directly from the observed data. Let $\underline{S}_E$ represent the residual or errors sums of squares and cross products matrix, and we obtain the variance-covariance estimate by dividing by the degrees of freedom scalar, N-q

(69) $$\hat{\underline{\Sigma}} = \frac{1}{N-q}\underline{S}_E = \frac{1}{N-q}\mathrm{Tr}(\hat{\underline{E}}'\hat{\underline{E}}) \qquad \text{from (64)}$$

$$= \frac{(\underline{Y} - \underline{X}\hat{\underline{B}})'(\underline{Y} - \underline{X}\hat{\underline{B}})}{N-q} \qquad \text{substituting}$$

$$= \frac{\underline{Y}'\underline{Y} - \hat{\underline{B}}'\underline{X}'\underline{Y}}{N-q} \qquad \text{by multiplication and cancelling}$$

or

$$= \frac{\underline{Y}'\underline{Y} - \hat{\underline{B}}'\underline{X}'\underline{X}\hat{\underline{B}}}{N-q} \qquad \text{a frequently encountered equivalent form}$$

It should be obvious that each element of $\hat{\underline{\Sigma}}$ is an estimate of population variance and covariance, with σ_j^2 on the diagonal and σ_{ij} $(i \neq j)$ on the off-diagonal cells.

As in the univariate case, the variance in $\underset{N\times p}{\underline{Y}}$ can be partitioned. Let $\underline{S}_Y$ represent the total sums of squares and cross products in the sample criterion scores, $\underline{Y}$, and $\underline{S}_R$ the variation in the predicted scores, $\hat{\underline{Y}}'\hat{\underline{Y}} = \hat{\underline{B}}'\underline{X}'\underline{X}\hat{\underline{B}}$, called the sums of squares for regression. Then the partition is

(70) $$\underline{S}_Y = \underline{S}_R + \underline{S}_E$$

which breaks the total variation into two components: that attributable to regression and that attributable to error. If $\underline{S}_R$ approximates $\underline{S}_Y$, then $\underline{S}_E$ will be small and the predictor scores will predict the criterion scores quite well. If $\underline{S}_R$ is small compared to $\underline{S}_Y$, then the linear model does not fit the data well and $\underline{S}_E$ will be quite large.

The "goodness of fit" of the model can be determined in a way that is identical to that in univariate regression. As Press (1972) indicates, "In multivariate regression, the value of R^2 can be computed for each equation separately to study the effectiveness of each relationship in accounting for observed variation" (p. 195).

As Timm (1975) indicates, "Estimation theory, using the multivariate linear model, is no different from employing $\underline{p}$ univariate models. It is not until hypothesis testing theory is employed that the models really differ. Univariate analysis does not address itself to the dependency that exists among a set of $\underline{p}$ response [criterion] variables" (p. 309). We turn now to the question of hypothesis testing.

Hypothesis Testing for Multivariate Regression

The logic of hypothesis testing in multivariate multiple regression is similar to that in its univariate counterpart except

that multiple criterion variables are tested simultaneously. The purpose of the simultaneous test is to maintain protection levels at a point predetermined by the researcher, i.e., at α. Were separate tests run for each dependent variable, the likelihood of obtaining a false significant value would increase in direct proportion to the number of dependent variables being tested, i.e., the power of the test decreases.

The general test for all coefficients of the raw form of the model is

$$H_o : \underline{B} = \underline{0} \tag{71}$$

which can be examined by using a multivariate analysis of variance (MANOVA) table. To conduct the test we must obtain the sums of squares and cross products matrices for both hypothesis and error, from which mean squares may be determined. These are, by extension from the univariate case

$$\underline{Q}_e = \underline{Y}'\underline{Y} - \hat{\underline{B}}\underline{X}'\underline{X}\underline{B} \tag{72}$$

and

$$\underline{Q}_h = \hat{\underline{B}}'\underline{X}'\underline{X}\hat{\underline{B}}$$

The MANOVA table can then be constructed

Insert Table 4 here

Normally, of course, we are not interested in testing all of the regression equations, since we seldom test the intercept and usually are interested in testing only a subset of the slope coefficients. Hence we would partition the model into two separate regressions and develop an hypothesis which would enable us to test the subset. The partitioned model would be

$$(73) \qquad \underline{Y} = \left[\underline{X}_I \; \underline{X}_{II}\right] \begin{bmatrix} \underline{B}_I \\ \underline{B}_{II} \end{bmatrix} + \underline{E}$$

with the hypothesis that

$$(74) \qquad H_o : \underline{B}_{II} = \underline{0}$$

Equations (73 and 74) simply assert that the first set of variables $\underline{X}_I$ (I = 1 . . . m) and regressions coefficients $\underline{B}_I$ (I = 0 . . . M) are not being tested while the remaining variables $\underline{X}_{II}$ (II = m . . . k) and coefficients $\underline{B}_{II}$ (II = m . . . k) are under test. It is, of course, possible to test the final set of coefficients, $\underline{B}_k$, by setting the indicator m to k-1. Then we would have $\underline{B}_I$ equal to all k-1 coefficients and $\underline{B}_{II}$ equal to the last set of k coefficients.

There are a number of alternative procedures available to test the hypothesis stated in equations (71 and 74). These all depend on solutions for the roots or eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_s$ of the characteristic equation

$$(75) \qquad |\underline{Q}_h - \lambda \underline{Q}_e| = 0$$

where the roots are ordered from largest, λ_1, to smallest, λ_s. Anderson (1958, Cp. 8) and Timm (1975, pp. 137-140, 146-149, and 308-313) describe a number of criteria which may be used to test the hypotheses. These tests are provided in Table 5. Tables for each of these distributions at selected degrees of freedom are available in Timm (1975). Since most of these distributions are unfamiliar to communication researchers, the table includes tests for multivariate hypotheses by approximations to χ^2 and F statistics which have been derived from the other distributions.

Table 5 about here

It is instructive to examine the logic of the Wilk's likelihood ratio criterion, Λ. It was mentioned earlier that the determinant of a variance-covariance matrix can be considered a measure of generalized variance. Wilk's Λ is the ratio of two generalized variances. The numerator contains the variance for the reduced model, Q_e, i.e., those variables not under test; the denominator contains the variance for the full set of predictors, i.e., those predictors not under test plus those that are being tested, Q_h. Thus

$$\Lambda = \frac{|Q_e|}{|Q_e + Q_h|} \tag{76}$$

It should be apparent that if $B_{II} = 0$ then $Q_h = 0$ and $\Lambda = \frac{|Q_e|}{|Q_e|} = 1$,

which is the upper bound of the statistic. On the other hand, to the extent that Q_h (i.e., X_m . . . X_k) adds variance over and above the reduced model, the denominator will increase and $\Lambda < 1$. Thus, the smaller Λ is, the more that Q_h or B is adding to criterion variation. The F and χ^2 tests simply transform the Λ statistic (really, the U distribution) to the more familiar F and χ^2 distributions. Most computerized MMR routines will print both Λ and either F or χ^2 test statistics.

Testing the criterion variables. If Λ (or any of the other test criterion) is significant it is then possible to determine which of the criterion variables are being affected by the variables in the predictor set. Univariate F-ratios may be used to accomplish this task. The univariate F that is largest is the one which shows which of the criterion variables is most affected by the predictors, and so on down to the smallest F-ratio which shows the criterion variable which is least affected by predictors. However, as Finn (1975, p. 156) points out, these tests are not independent: "there is no necessary relationship of the significance of the univariate and multivariate tests for one hypothesis. For example, one or more univariate F's may be significant and not the multivariate statistic, or vice versa" (p. 157). Consequently, it is usually recommended that in addition to the univariate F tests, researchers examine the simple and multiple correlations to aid in interpretation.

To deal with this problem of interdependence of multivariate and univariate tests for locating effects, it is possible to conduct a step-down analysis which will provide independent univariate tests for the criterion variables. The term step-down as used here and by Bock (1966) and Finn (1975, pp. 157-160) refers to the fact that the criterion variables rather than the predictor variables are being examined (eliminated) sequentially. The tests, which are described by Roy (1958) require that the researcher impose an a priori ordering on the criterion variables. If the researcher has no rational basis for such an ordering, then step-down tests will be of little value. The procedures for conducting the tests are identical to calculating F tests (for each of the p criterion variables) with the condition imposed that only the conditional variance in the criterion variables be analyzed. The tests proceed by regressing all the q predictor variables on y_1, on y_2 eliminating y_1, on y_3 eliminating y_1 and y_2, etc. The test procedure is a form of backward elimination. First, the p^{th} criterion variable y_p is eliminated, i.e., the last one in the ordered set. Then the next to the last is eliminated y_{p-1}, etc., down to the first criterion, y_1. At each step an F statistic is calculated; if F is non-significant, and H_o for a particular criterion cannot be rejected, it means that the predictor set does not significantly contribute to variation in particular criterion variable y_j. Each time that

H_0 cannot be rejected, the step-down procedure continues until such a time as an H_0 must be rejected in favor of the alternative H_A. At that point testing terminates since all remaining tests are nonindependent. See Bock (1966) and Finn (1974) for additional details on step-down analysis.

Testing the predictor variables. Having determined how the set of predictor variables differentially affects the criterion variables it is often of interest, as in the univariate case, to test hypotheses regarding a subset of the predictor variables. Tests for contributions to criterion variance of individual predictors are facilitated by creating orthogonal columns of B, i.e., converting the partial regression coefficients to semipartial regression coefficients. Each semipartial regression coefficient accounts for variance in Y not accounted for by predictor variables that preceed it in the regression equation. This orthogonalization process, of course, requires that the researcher impose an a priori order on the predictors. When the backward sequential tests on the predictor variables encounter a variable which significantly contributes to criterion variance, the testing procedure must stop, since all subsequent tests are non-independent. It should be apparent from this discussion that the researcher must rely heavily upon his theory to specify the order of the predictor variables so that the strongest known predictor is the first variable in the equation and the weakest is last.

Reestimation

As in the univariate case, if a researcher either adds or deletes variables from the equation with which he began, it is important to reestimate the parameters of the resultant or final equation. Variance in the non-significant variables is pooled with error variance to make the final estimate. Finn (1975) indicates that "Under ideal circumstances, these final estimates should be obtained from a sample other than the one used for significance tests" (p. 165).

Issues and Implications

In this final section of the paper we shall briefly explore several topics which are raised by or related to MMR. These should help to place the information just presented in a broader context.

Algorithms for Selecting Predictor Variables

There are a number of computerized algorithms which can be used to select an optimal set of predictors for a regression equation. In fact, Draper and Smith (1966) identify six different alternative procedures. Four of these are summarized in this section; for details the reader is referred to the fuller explication in Draper and Smith and also to Kerlinger and Pedhazur (1973, pp. 285-295).

The first procedure is termed all possible regressions. Here separate regression equations are calculated for all possible combinations of predictor variables: (1) each predictor variable

examined separately, (2) all possible predictor pairs, (3) all possible triplets, (4) and in general, all possible k-tuples (i = 1 . . . k predictors), where the final k-tuple contains one set of all k variables. A moments reflection will indicate that a total of 2^k separate regressions must be determined since each variable may either be included or excluded from the equation. Within each of the k-tuple regressions (i.e., pairs, triplets, etc.), equations are ordered according to some criterion, usually the amount of variance accounted for, R^2. The researcher then selects what he considers the best equation, often by looking for "any consistent pattern of variables in the leading equations in each set" (Draper and Smith, p. 162).

The second procedure is called the backward elimination. There are three basic steps. First, calculate a regression equation which includes all the k predictor variables. Second, calculate partial F-tests for each of the k variables which indicates the significance of X_k as if it were the last variable entered into the equation. Finally, select the lowest partial F value and compare it with a partial-F value set equal to some predetermined level of significance, α. If the smallest partial-F is less than F_α, then delete that variable and repeat the process for k-1 predictor variables. This sequence continues until the smallest partial-F at any given step (k-1 predictor variables to 1 predictor) is greater than F_α. The

variables that remain are considered significant predictors.

The third alternative for choosing predictors is forward selection. In this process variables are added to the regression equation one at a time rather than deleted from the equation as was done in backward elimination. The first step is to select the largest zero-order simple correlation coefficient between y and X_j. This variable is used to construct a simple regression equation. Second, an F-test is calculated to determine whether the simple regression of y on the first predictor variable is significant. Assuming this test is significant, the third step is to calculate first-order partial rs between y and all other X_j controlling for the predictor variable already in the simple regression equation. The X_j with the highest partial r (or squared partial r) is then selected for the construction of a multiple regression equation with two predictor variables. Fourth, as with the backward elimination procedure, partial F-tests are undertaken to determine whether the variable selected for inclusion contributes significantly to the variance in y over and above the variance accounted for by those predictors already in the equation, i.e., R^2 for the simple regression and R^2 for the two variable multiple regression. Finally, this process (calculating higher order partial rs, selecting the one of greatest magnitude, creating a new regression equation with this variable added, calculating the significance of

the increase in variance contributed by this variable) continues until the partial-F test indicates that the variable selected for inclusion at the next iteration is non-significant. The final equation includes all variables found significant up to the point at which adding an additional predictor variable does not significantly increase the amount of variance accounted for in the criterion variable; all remaining predictors are considered non-significant and omitted from the equation.

The fourth procedure, which is called stepwise regression, is really a variation on the forward selection procedure. For any given set of variables, the magnitude of regression coefficients does not vary according to their order in a regression equation. For example, the regression coefficients for three variables, call them A, B, and C, will be the same regardless of the order in which these three variables are arranged in a 3 variable multiple regression equation, i.e., whether variable A is first, second, or last, whether variable B is first, second . . ., etc. If, however, a variable, say D, is added to this equation, or one of the three variables, A, B, or C, is deleted from the equation, it is highly likely that the magnitude of all the coefficients in the new (4 variable or 2 variable) equation will be considerably altered. The reason for this is that regression coefficients are a patterned function of a set of variables, such that they account for as much

variance as possible in $\underline{y}$ or $\underline{Y}$, given that particular set of predictors and that set only. The amount that the regression coefficients will change is a function of the degree to which they share variance with each other and the criterion variable(s); the more variance they share in common, the greater the coefficients will change by adding or deleting variables. It should be apparent that the three procedures identified so far all require the addition or deletion of variables. The stepwise procedure attempts to compensate for the fact that at each step adding a variable (remember, it is a variation on the forward selection procedure) could reduce the amount of variation contributed by one or more of the variables already in the equation to a point at which it (they) would no longer be considered significant. Thus, after each step in which a variable is selected for inclusion in the next larger multiple regression equation, a partial-$\underline{F}$ test is performed on all variables in the equation to which the new variable is being added to determine whether each still significantly contributes to criterion variance. Any that fail to meet $\underline{F}_{\alpha}$ are deleted from the regression equation and join the set of predictor variables not in the equation. Then the process is repeated: new $n^{\underline{th}}$ order partial rs are calculated; the largest is selected for inclusion in a new regression equation; a test of the significance of this new variable is conducted; if it exceeds $\underline{F}_{\alpha}$ for appropriate $\underline{\nu}$, a

separate partial-F is conducted for each criterion variable already in the equation; any criterion that fails the F_{α} is removed from the equation. The process terminates when there is no new variable that can be added to the equation and significantly increase criterion variation.

An alternative procedure, called stagewise regression will not be discussed since it is not a least squares procedure. Also, it should be noted that there are some "variations of the themes" presented in these four alternatives (see Draper and Smith, 1966, pp. 172-173).

Given the four procedures just described. which one should the researcher choose? That, of course, is a judgment call, which each person will have to make depending upon his research needs. In general, however, all possible regressions is both impractical and without recourse to statistical test; it should be avoided. Draper and Smith (p. 172) recommend the stepwise procedure but indicate that it can be abused by the inexperienced. In this context, Finn (1974), makes an additional important point:

> "Stepwise" procedures which attempt all possible orderings, or search for the best single prediction equation do not generally yield valid test statistics, and must be interpreted with caution. With a predetermined order of predictor variables, valid sequential test statistics are obtained. Using a fixed order, it is

also possible to test important combinations or sets of variables (p. 161).

Testing the Assumptions

In this paper we have stressed the importance of the assumptions of the model being employed. The only way to determine whether a given set of assumptions is viable for any given study is to test them. As Timm (1975) indicates, "Plots of residuals against the fitted values, against the independent variables, and sometimes against variables not included in the model help to determine (1) whether model assumptions are reasonable, (2) the linearity of the regression function, and (3) whether important variables have been left out of the model" (p. 269). Draper and Smith (1966) review these procedures in detail and Daniel and Wood (1971) give several illustrative examples; the procedures will be discussed only briefly here.

The logic of the examination is as follows: Recall that our assumptions for the classical normal linear regression model specified that the errors were normally distributed with zero mean and constant variance and that they were independent. If the assumptions are correct, the errors or residuals ought to display these characteristics.

The first procedure is to undertake an overall plot of the residuals. Normality, or departure from normality, can be judged by reference to a table of random normal deviates or by ploting

the residuals on standard probability paper. Alternatively, the residuals may be transformed to "unit normal deviate" form, in which case we would expect that 95% of the residuals would fall within the $\pm$ 1.96 σ limits. Outliers, which are residuals which lie far out in the tails of the distribution, say $\pm$ 3 or 4 standard deviations, can also be identified. Some standard regression computer packages now have subroutines which permit the plotting of residuals in unit normal deviate form.

Another procedure is to plot each residual against the predicted value which helped to generate it, i.e., $\hat{Y}_i$. Here, a horizontal band of scores lieing relatively equidistant from zero would indicate that the assumptions have been met. If the distribution of residuals shows divergence (or convergence) across the $\hat{Y}_i$ scores, the constant variance assumption is challenged. If the variance is constant, but there is an upward or downward trend in the plot, then there has probably been an error in the analysis or the constant term, β_0, has been omitted. Finally, if the plot looks curvilinear, then the assumption of linearity is questioned. Transformation on the variables or extra terms (square or cross-products) may be required in the model.

Residuals may be plotted against the independent variables. As with the plot against $\hat{Y}_i$, the existence of a horizontal band of residuals is the desired form. Failure to obtain such a plot would raise similar sorts of questions as those in the previous paragraph.

Statistics are available for formally testing residuals. The reader is referred to the work by Anscombe (1961) and Anscombe and Tukey (1963) for the details. Other plots are also possible and Draper and Smith (1966) recommend that any residual plot should be made that makes sense in light of the research (e.g., plotting residuals against a variable measured on the same subjects but not included in the regression equation). For an example of testing assumptions in communication research see Monge, et al. (1976), who examined the plot of the residuals for a fully recursive structural equation model of the determinants of communication structure in large organizations.

Residuals, of course, are not the only way to test assumptions. Autoregression can be ascertained by the Durbin-Watson test. Multicollinearity can be determined by examining the determinant of ($\underline{X}'\underline{X}$); if it approaches zero some column of the matrix is linearly dependent upon some other column(s) of the matrix. Homoskedasticity can be examined by a homogeneity of variance test. Other procedures are also available and can be found in most standard econometrics texts (e.g., Goldberger, 1964; Johnston, 1972, Kmenta, 1971). The important point to remember from this section is that when communication researchers undertake a regression analysis, they should always test the assumptions of the model. Only when this becomes standard practice in communication research will we be able to have confidence

in research findings that employ statistical inference.

Comparisons with Related Techniques

MMR has been presented in the context of its univariate counterpart, linear multiple regression. To provide a broader context it might be useful to examine the relationship between MMR and several other techniques. I will briefly discuss: (1) the general linear model, (2) MANOVA, (3) canonical correlation, and (4) general systems of equations.

MMR has been treated as a special case of the general linear model which consists of the functional form of the model and the assumptions specified in the Gauss-Markoff theorem (set-up). Many of the multivariate techniques, e.g., MANOVA, canonical correlation, etc., are derivable from the general linear model. To understand that point provides two important insights. First, it emphasizes that many of the multivariate statistics are highly interrelated and not separate, independent techniques. Second, it emphasizes how the various techniques differ as a function of the differences in their assumptions, so that modifying an assumption of the general linear model makes it necessary to choose an alternative form of the model (i.e., a different multivariate technique).

Comparing MMR with MANOVA helps to illustrate these two points. First, Timm (1975) demonstrates how both are special cases of the multivariate general linear model. Second, if the

researcher assumes categorical rather than continuous measurement on X_j then MANOVA is the appropriate form of the model to utilize. Actually, as Bochner and Fitzpatrick (1977) illustrate in another paper for this conference, a MANOVA model can be analyzed with MMR techniques by use of special (dummy) coding techniques (See also Kerlinger and Pedhazur, 1973; Press, 1972; Kmenta, 1971).

Canonical correlation is a technique that is appropriate when there are two sets of variables, measured on ordinal scales. The coefficient, R_c is the simple correlation between two random variables which are each linear composites of two or more variates. Each composite is defined by weights applied to its variates designed to maximize its correlation, R_c, with the other composite. As Finn (1974) indicates, when one of the "composites" consists of only one variate, R_c becomes "the multiple correlation of one measure with the other set. The weights are the partial regression coefficients" (p. 188). MMR is the appropriate technique when the researcher is interested in explaining variation in one set of variables by variation in the other; canonical correlation is appropriate when one is interested in forming composites which maximally share variance with each other.

Some econometrics texts treat MMR under the topic of sets of linear regression relations (See, e.g., Goldberger, 1964, pp. 201-212). This extension to sets of dependent variables from the univariate case is, as we have shown, quite straightforward. Yet always the

emphasis is on explaining variance in the criterion variables by variation in the predictor variables. If, however, a researcher is interested in testing a theory which consists of a system of linear relations (either single equations which are part of a larger system of equations or the entire system itself), then the OLS estimation procedures described in this paper are appropriate only in the special case of a fully recursive system. Simultaneous linear structural equation systems require alternative estimation procedures which are the topic of other papers in this conference (Cappella, 1977; Fink, 1977).

Coping with Failed Assumptions

In this paper considerable emphasis has been placed on the importance of the assumptions in the classical linear model. Our discussion, however, has been limited to OLS techniques; we have not examined alternative estimation procedures. Though it is beyond the scope of this paper it is important to point out that procedures have been developed to permit statistical inference in spite of violations of OLS assumptions. Some of these techniques, such as two stage least squares to estimate autoregressive disturbances and the use of instrumental variables for errors in variables (i.e., measurement error in X), are simply more sophisticated applications of OLS procedures. Others require complete abandonment of OLS in favor of maximum likelihood

estimators. Whatever the case may be, the communication researcher is strongly encouraged to consult multivariate and econometric texts to discover the appropriate alternatives which are necessitated by violations of assumptions in his or her data.

Advantages and Disadvantages of MMR

Kerlinger and Pedhazur (1973) review in depth a number of social science studies that employed multiple regression; their comments give the reader an excellent notion of the breadth of applications that are possible with the technique (See Cps. 15 and 16). They also discuss the limitations and strengths of regression techniques (pp. 441-445) and that material, some of which has already been mentioned, will not be repeated here. Rather, I would prefer to make three brief points.

First, MMR techniques lend themselves to simultaneously replicating previous research and developing new findings. This can occur by including predictors in the regression equation which have been shown to be important in earlier research. New variables may then be added and tested for significance over and above those already in the equation. In this way results become cumulative and we can build communication theories of known predictors.

The second point is that regression analysis, if done correctly, virtually necessitates the formal specification of the regression equation. To my way of thinking this particularly facilitates

interpretation of complex relations, particularly interaction terms. Furthermore, since traditional ANOVA and more complex MANOVA designs may be analyzed via regression techniques, it is possible to develop "mixed" designs which include both categorical and continuous predictors.

Third, though we have only discussed cross-sectional data in this paper, all econometrics texts also discuss estimation procedures for time series data. These may vary from simple lagged variables to complex simultaneous structural equation systems. These procedures can be used to explicitly capture the time-variant, processional nature of many communication phenomena.

Conclusion

At the outset of this paper we began with a quotation from Tukey (1962) which asserts not only that our most important questions are more likely to be vague than precise, but also that our best answers will often be approximate rather than exact. Asking the right questions is a theoretical endeavor; offering our best approximations to answers is a statistical undertaking.

Having shown in this paper how analysts may expand the relations they can examine via regression techniques to include multiple criterion variables, it seems important to stress the necessity for theorists to develop formulations which will incorporate this expanded capability. To develop multivariate theories that can be

studied by multivariate multiple regression should lead us a long way toward asking the right questions and obtaining regression coefficient approximations to laws of human communication.

References

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.

Anscombe, F. J. Examination of residuals. Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics, and Probability, 1961, I, 1-36.

Anscombe, F. J., & Tukey, J. W. The examination and analysis of residuals. Technometrics. 1963, 5, 141-160.

Blalock, H. M., Jr. Causal inferences in nonexperimental research. Chapel Hill: The University of North Carolina Press, 1964.

Blalock, H. M., Jr. (Ed.) Causal Models in the Social Sciences. New York: Aldine, 1971.

Bochner, A., & Fitzpatrick, M. Multivariate analysis of variance in communication research. Paper presented to the conference on Multivariate Techniques in Communication Research. Asilomar Conference Center, Pacific Grove, California, April 1977.

Bock, R. D. Contributions of multivariate experimental designs to educational research. In R. B. Cattell (Ed.), Handbook of multivariate experimental psychology. Skokie, Ill.: Rand McNally, 1966.

Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill, 1975.

Cappella, J. N. Structural equation modelling. Paper presented at the conference on Multivariate Techniques in Communication Research. Asilomar Conference Center, Pacific Grove, California, April 1977.

Daniel, C., & Wood, F. S. Fitting equations to data. New York: Wiley, 1971.

Draper, N. R., & Smith, H. Applied regression analysis. Wiley, 1966.

Fink, E. Structural equation modelling: Unobserved variables. Paper presented to the conference on Multivariate Techniques in Communication Research, Asilomar Conference Center, Pacific Grove, California, April 1977.

Finn, J. D. A general model for multivariate analysis. New York: Holt, 1974.

Goldberger, A. S. Econometric theory. New York: Wiley, 1964.

Johnston, J. Econometric methods (2nd ed.). New York: McGraw-Hill, 1972.

Kerlinger, F. N., & Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, 1973.

Kmenta, J. Elements of econometrics. New York: Macmillian, 1971.

Kshirsagar, A. M. Bartlett decomposition and Wishart distributions. Annals of Mathematical Statistics, 1959, 30, 239-241.

Monge, P. R., Boismier, J. M., Cook, A. L., Day, P. D., Edwards, J. A., & Kirste, K. K. Determinants of communication structure in large organizations. Paper presented to the Information Systems Division of the International Communication Association, Portland, Oregon, April 1976.

Press, S. J. Applied multivariate analysis. New York: Holt, 1972.

Roy, J. Step-down procedure in multivariate analysis. Annals of Mathematical Statistics, 1958, 29, 1177-1187.

Timm, N. H. Multivariate analysis with applications in education and psychology. Monterey, California: Brooks/Cole, 1975.

Tukey, J. W. Causation, regression, and path analysis. In Kempthorne, O., et al. (eds.), Statistics and Mathematics in Biology. Ames, Iowa: Iowa State College Press, 1954.

Tukey, J. W. The future of data analysis. Annals of Mathematical Statistics, 1962, 33, 1-67.

		Criterion Variable(s)	
		Univariate ($P = 1$)	Multivariate ($P \geq 2$)
Predictor Variable(s)	Simple ($K = 1$)	1. $\underset{NxM}{A} = \left[\underset{nx1}{y} \vdots \underset{nx1}{x}\right]$	3. $\underset{NxM}{A} = \left[\underset{NxP}{Y} \vdots \underset{Nx1}{x}\right]$
	Multiple ($K \geq 2$)	2. $\underset{NxM}{A} = \left[\underset{Nx1}{y} \vdots \underset{NxK}{X}\right]$	4. $\underset{NxM}{A} = \left[\underset{NxP}{Y} \vdots \underset{MxK}{X}\right]$

Table 1. Partition of a data matrix, $\underset{NxM}{A}$, into four possible combinations of Predictor Variables, $\underset{NxK}{X}$, and Criterion Variables, $\underset{NxP}{Y}$, submatrices, $M = P + K$. The four cells provide the data partitions for (1) univariate simple regression, (2) univariate multiple regression, (3) multivariate simple regression, and (4) multivariate multiple regression.

Table 2. ANOVA Table for Testing $\underline{\beta} = \underline{0}$.

(from Timm, 1975, p. 273)

Source	df	SS	E(MS)	F
Total Regression	$k + 1$	$Q_h = \hat{\underline{\beta}}'\underline{X}'\underline{X}\hat{\underline{\beta}}$	$\sigma^2 + \frac{\underline{\beta}'\underline{X}'\underline{X}\underline{\beta}}{k + 1}$	MS_h/MS_e
Residual	$N - k - 1$	$Q_\varepsilon = \underline{y}'\underline{y} - \hat{\underline{\beta}}'\underline{X}'\underline{X}\hat{\underline{\beta}}$	σ^2	
Total	N	$Q_T = \underline{y}'\underline{y}$		

Table 3

Hypotheses and Test Statistics for Testing Regression Coefficients

Hypothesis	Test Statistic	Comment
(1) $H_o : \hat{\beta}_k = 0$	$t = \frac{\hat{\beta}_k}{S_{b_k}}$	$\hat{\beta}_k$ equals zero. Can be used to test the intercept, β_o. See Johnston, 1972, p. 138.
(2) $H_o : \hat{\beta}_k = M_k$	$t = \frac{\hat{\beta}_k - M_k}{S_k}$	$\hat{\beta}_k$ equals some constant, M_k. See Kmenta, 1971, p. 366.
(3) $H_o : \hat{\gamma} = 0$ or $\hat{\beta}_1 \ldots \hat{\beta}_k = 0$	$F = \frac{SS_R/(k-1)}{SS_E/(N-k-1)} = \left(\frac{N-k}{k-1}\right)\left(\frac{R^2}{1-R^2}\right)$	All coefficients except the intercept are zero. See Goldberger, 1964, p. 176; Timm, 1975, p. 273.
(4) $H_o : \hat{\beta}_h = 0$	$F = \left(\frac{R_o^2 - R_h^2}{1-R_o^2}\right)\left(\frac{N-k-1}{k-m}\right)$	Some subset of the coefficients (perhaps the final one) is zero. See Goldberger, 1964.
(5) $H_o : \beta_j = \beta_k$ (for $j \neq k$)	$t = \frac{\hat{\beta}_j - \hat{\beta}_k}{S_{\hat{\beta}_j - \hat{\beta}_k}}$	One regression coefficient equals another. See Goldberger, 1964, p. 175.
(6) $H_o : \beta_j + \beta_k = a$	$t_{N-k} = \frac{\hat{\beta}_j + \hat{\beta}_k - a}{S_{\hat{\beta}_j - \hat{\beta}_k}}$	The sum of two (or more = k) coefficients equals a constant. See Kmenta, 1971, p. 372.
(7) $H_o : \hat{\beta} = \hat{\delta}$ ($\hat{\delta}$ is the vector of regression coefficients for a second equation)		Two regression equations are equal. See Kmenta, 1971, p. 373 for the test statistics.

Table 4. MANOVA Table for Testing $\underline{B} = \underline{0}$

From Timm, 1975, p. 309.

Source	df	SS	E(MS)
Total Regression	$k + 1$	$Q_h = \hat{\underline{B}}'\underline{X}'\underline{X}\hat{\underline{B}}$	$\underline{\Sigma} + \frac{\underline{B}'\underline{X}'\underline{X}\underline{B}}{k + 1}$
Residual	$N - k - 1$	$Q_e = \underline{Y}'\underline{Y} - \hat{\underline{B}}'\underline{X}'\underline{X}\hat{\underline{B}}$	$\underline{\Sigma}$
Total	N	$Q_T = \underline{Y}'\underline{Y}$	

Table 5

Criteria and Tests for Multivariate Hypotheses

$$H_o : \underline{B} = \underline{0};\ H_o : \underline{\Gamma} = \underline{0};\ H_o : \underline{B}_{II} = \underline{0}$$

1. Wilks' likelihood ratio criterion

$$\Lambda = \frac{|\underline{Q}_e|}{|\underline{Q}_e + \underline{Q}_h|} = \sum_{i=1}^{s} (1 + \lambda_i)^{-1} < U^{\alpha}(P,\ k + 1,\ N - k - 1)$$

2. Roy's largest root criterion

$$\theta_s = \frac{\lambda_1}{1 + \lambda_1} > \theta^{\alpha}\ (s,\ m,\ n)$$

3. Lawley-Hotelling trace criterion (or Hotelling's generalized T^2 statistic)

$$U^s = \frac{T_o^2}{N-k-1} = \sum_{i=1}^{s} \lambda_i > U_o^{\alpha}\ \ (s, m, n)$$

4. Pillai's trace criterion

$$V^s = \sum_{i=1}^{s} \frac{\lambda_i}{1 + \lambda_i} > V^{\alpha}(s, m, n)$$

5. Bartlett's χ_B^2 test

$$\chi_B^2 = -\left[(N-1) - \tfrac{1}{2}(P + k + 1)\right] \text{Log } \Lambda > \chi_{\alpha}^2\ (pk)$$

6. Fisher's F test

$$F = \frac{\nu_e}{\nu_h} \cdot \frac{1 - \Lambda}{\Lambda} > F^{\alpha}\ (\nu_e,\ \nu_h)$$

where

$s = \min\ (\nu_h,\ p) = (k+1,\ p)$

$m = \dfrac{|U - \nu_h| - 1}{2}$

$n = \dfrac{\nu_e - U - 1}{2}$

ν = degrees of freedom